

Latent profiles of vocabulary and domain knowledge and their relation to listening comprehension in kindergarten

HyeJin Hwang 

Florida Center for Reading Research, Florida State University, Tallahassee, Florida, USA

Sonia Q. Cabell 

Florida Center for Reading Research, Florida State University, Tallahassee, Florida, USA

Background: Vocabulary and domain knowledge are important factors that influence comprehension development. However, these factors have most frequently been examined in relation to reading comprehension and much less frequently examined in relation to listening comprehension. Moreover, almost no empirical studies have examined profiles of strengths and weaknesses in vocabulary and domain knowledge in students at the beginning of schooling. In the current study, we investigated different profiles (groups) of students regarding receptive and expressive vocabulary and science and social studies domain knowledge in the fall of kindergarten, as well as relations between the profiles and listening comprehension development throughout kindergarten.

Method: We analysed data collected from 629 kindergarten students living in two large urban districts in the United States by conducting a series of latent profile and chi-square analyses.

Results: Latent profile analysis identified three groups of students that displayed distinct patterns of achievement in vocabulary and domain knowledge, namely, students whose scores on the four measures were below average (-1.6 standard deviations [SD]), average (-0.2 SD) and above average (0.8 SD). There was no group of students who displayed unbalanced achievement in vocabulary and domain knowledge at the beginning of schooling. In addition, statistically significant differences in listening comprehension development during kindergarten were observed for all pairwise comparisons across the groups.

Conclusion: The findings appear to indicate that students can acquire vocabulary to the extent that they develop domain knowledge and vice versa. The commensurate achievement might be due to mutually enhancing relations between vocabulary and domain knowledge. Moderate to large effect sizes on listening comprehension development when comparing the groups suggest the importance of supporting vocabulary and domain knowledge from the early grades.

Keywords: vocabulary, domain knowledge, listening comprehension, latent profile analysis

Highlights

What is already known about this topic

- Prior studies have shown that vocabulary and domain knowledge predict comprehension, mostly studied in relation to reading comprehension.
- Early listening comprehension predicts later reading comprehension.

What this paper adds

- This is one of the first person-centred studies that reveals how vocabulary and domain knowledge manifest in students at the beginning of schooling (kindergarten).
- The study extends a body of literature on comprehension by focusing on listening comprehension, jointly predicted by profiles of vocabulary and domain knowledge.

Implications for theory, policy or practice

- Commensurate achievement in vocabulary and domain knowledge might indicate the mutually enhancing relation between the two in early schooling.
- Literacy instruction is recommended to support vocabulary and domain knowledge simultaneously (e.g., explicit teaching of vocabulary during content-rich read-alouds).

Proficient listening comprehension is crucial for academic advancement, particularly at the beginning of schooling when most class instruction and activities are mediated via oral communication. Students who understand spoken language well in the early grades are also likely to develop strong reading comprehension in later grades (Storch & Whitehurst, 2002). Although proficient listening comprehension is fundamental for academic achievement, there is considerable variability among students in the development of listening comprehension from the early childhood years (Hagen et al., 2017). Examining what contributes to individual differences in listening comprehension can help us better understand and support listening comprehension.

Development of listening comprehension is a complicated process that involves multiple components (Nation et al., 2010). Among them, vocabulary plays an important role. The more students know the meanings of words, the more accurate inferences they can generate from oral messages, resulting in better listening comprehension (Kim, 2020). Another important, but less-studied, factor of listening comprehension is domain knowledge (i.e., broad prior knowledge related to a field of a study). Strong domain knowledge may facilitate inference generation on missing information in a text, a crucial process of comprehension (Kintsch, 2013).

Investigating strengths and weaknesses (i.e., profiles) of vocabulary and domain knowledge can reveal potential sources of difficulties in listening comprehension. However, little is known about profiles of vocabulary and domain knowledge when students are in the early primary grades. It is possible that students would display comparable achievement in vocabulary and domain knowledge as some regard vocabulary as part of domain knowledge (Cook & Gueraud, 2005). Alternatively, there may be a group of students who know many words but have relatively weak domain knowledge (e.g., a child

knows the meaning of *fruit* without knowing how it relates to plants), as well as another group of students who have strong domain knowledge with relatively weak vocabulary (e.g., a child knows some trees have green needle-like leaves without knowing the word, *evergreen*).

Moreover, students with different profiles of vocabulary and domain knowledge may display different listening comprehension development. For example, a hypothetical group of students with high vocabulary and low domain knowledge may demonstrate more improvement in listening comprehension over time, when compared with another hypothetical group of students with low vocabulary and high domain knowledge. One reason for this might be that students in the early grades are not as efficient in leveraging their domain knowledge to comprehend (e.g., Brandão & Oakhill, 2005) and tend to rely more on words they know. Thus, the difference in vocabulary between the two groups may contribute to the difference in listening comprehension. It appears, however, that no empirical study has investigated this matter.

Therefore, the aims of the current study are to (1) identify whether there are distinct profiles of students at the beginning of kindergarten who display similar strengths and weaknesses in vocabulary and domain knowledge by using latent profile analysis; and (2) examine differences in listening comprehension development among the profiles throughout the kindergarten year. We focused on the kindergarten year to understand profiles of students at the beginning of schooling when most instruction is mediated orally.

Comprehension, vocabulary and domain knowledge

The construction-integration model of comprehension. Comprehension is the process of constructing a mental representation of spoken or written text by extracting what text says (the textbase) and integrating the textbase with what listeners or readers already know (situation model of the text; Kintsch, 1998). To build the textbase, listeners or readers need to extract idea units from the text and make coherent connections among the idea units. Knowing many words in a text can facilitate meaning extraction of idea units from individual phrases or sentences (Anderson & Freebody, 1981). Establishing coherent connections among idea units depends on inference making, which is supported by vocabulary and domain knowledge. Efficient access to word meanings can free up attentional resources necessary to make inferences about how individual ideas are connected to one another (Mezynski, 1983). Domain knowledge can facilitate the generation of gap-filling inferences about missing information in the text, which can bolster the process of combining idea units (Elbro & Buch-Iversen, 2013).

Whereas the facilitative role of vocabulary in comprehension is related to extracting and connecting individual ideas, the role of domain knowledge extends beyond building the microstructure (Stahl et al., 1991). Domain knowledge aids in determining the relative importance of ideas (i.e., macrostructure of the textbase) and help listeners or readers direct their focus (e.g., focusing more on story elements than details of the story). Moreover, in order to construct the situation model, domain knowledge needs to be integrated with the textbase, which is a crucial process to gain new knowledge from interacting with the text (Kintsch, 1998).

As comprehension involves complex processes, it is important to support students' comprehension development from early grades, even before and while students are developing decoding skills. Strengthening listening comprehension in early grades can help prevent

difficulties in reading comprehension in later grades (Duke & Carlisle, 2011), as the texts that students must comprehend in later grades become increasingly complex (Hogan et al., 2014). Thus, the current study focused on listening comprehension development during the kindergarten year. Because listening comprehension develops rapidly during the kindergarten year (Hagen et al., 2017) and vocabulary and knowledge may jointly influence this development, we focused on the developmental change during the year, rather than examining concurrent relations between profiles and listening comprehension at a single time-point.

Vocabulary and comprehension. Consistent with Kintsch's theory, empirical studies have established the positive role of vocabulary in comprehension development (e.g., Wright & Cervetti, 2016). Vocabulary is related to listening (e.g., Sénéchal et al., 2006; Ricketts et al., 2007) and reading comprehension (e.g., Richter et al., 2013) in the elementary years. Given the important role that vocabulary plays in comprehension development, a more nuanced understanding is needed of the development of vocabulary in the early grades.

Vocabulary is commonly represented in two modalities: receptive and expressive. Receptive vocabulary is assessed by asking students to identify a picture or object that indicates the word provided by the assessor, while expressive vocabulary is assessed by asking students to provide a word that matches a presented picture or object. The two modalities are highly correlated, but instructional support during preschool can facilitate differential growth in receptive and expressive vocabulary prior to school entry (Sénéchal, 1997). For example, Hargrave and Sénéchal (2000) found that preschoolers' expressive vocabulary improved when they were engaged in dialogic book reading but not receptive vocabulary (see also Lonigan & Whitehurst, 1998). Gonzalez et al. (2014) demonstrated that conversations between teachers and preschoolers after read-alouds predicted expressive vocabulary, whereas teachers' talk regarding relations among words and concepts during read-alouds predicted receptive vocabulary. It is possible, then, that students might start the kindergarten year with different levels of receptive and expressive vocabulary because of different instructional support during preschool.

Domain knowledge and comprehension. Domain knowledge is a broad operationalisation of prior knowledge as it comprehensively includes knowledge related to a field of study (Cervetti & Wright, 2020). In the school context, domain knowledge is often considered broad knowledge in the science or social studies content areas (e.g., Tarchi, 2010) and has been studied in relation to comprehension in general (rather than comprehension on a specific topic). There has been much less research on the role of domain knowledge in comprehension in the elementary years, compared with *topic knowledge*, which is a narrow operationalisation of prior knowledge specifically related to the topic of a text being comprehended (e.g., plant survival; Hwang, 2020). Previous studies demonstrating the positive role of prior knowledge have mostly used measures of topic knowledge to understand comprehension on a specific topic (Hwang & Duke, 2020).

An emerging body of correlational studies has indicated that domain knowledge can predict reading comprehension in the elementary years (e.g., Hwang & Duke, 2020). However, very little attention has been paid to understanding the role of domain knowledge in comprehension development in early years of schooling. Based on the findings of previous studies with students in upper grades (e.g., Hwang, 2019), it is reasonable to surmise that domain knowledge would also play a facilitative role in comprehension development in early years of schooling. However, it is also plausible that in early years, domain

knowledge may not contribute to explaining the variance of comprehension because students in early grades often have difficulties in drawing on what they know to make inferences (Cain et al., 2001). For example, Brandão and Oakhill (2005) demonstrated that 7- or 8-year-old students had difficulty in integrating information in text with relevant information from their knowledge, after they analysed students' incorrect answers to reading comprehension questions. As there is still uncertainty on this matter, investigation on the role of domain knowledge in comprehension development in early years of schooling is needed. In early years of schooling, comprehension can be more accurately measured with listening comprehension than reading comprehension because students are still developing decoding skills (Lonigan & Burgess, 2017).

Moreover, as most studies have focused on either science or social studies domain knowledge to understand reading comprehension, rather than both domains simultaneously (e.g., Hwang, 2019; Hwang & Duke, 2020), there is a gap in a body of literature about students' knowledge development in different domains. Children can and are expected to develop understanding of different topics in science and social studies from early ages (Gonzalez et al., 2011), but varied exposure to and experiences in learning about science and social studies during pre-kindergarten (e.g., Chien et al., 2010) can lead students to develop different levels of science and social studies domain knowledge at the beginning of schooling. As the development of domain knowledge is characterised as a cumulative process (Neuman & Roskos, 2012) that can influence subsequent learning of the natural and social world, it is important to understand the strengths and weaknesses in science and social studies domain knowledge at the beginning of schooling.

It appears that no research has examined how vocabulary and domain knowledge jointly predict the development of listening comprehension. Most similar studies have used a variable-centred approach (e.g., regression) to examine how vocabulary and topic knowledge simultaneously relate to reading comprehension on that topic (e.g., Rydland et al., 2012). The variable-centred approach is limited in examining heterogeneity in a population regarding a given phenomenon. Therefore, we used latent profile analysis (LPA), a person-centred approach, to identify mutually exclusive groups, each comprising individuals with similar characteristics (see Nylund-Gibson et al., 2019). That is, LPA was utilised to explore the heterogeneity of receptive and expressive vocabulary and science and social studies domain knowledge across kindergarteners, which can provide a nuanced understanding of the joint contribution of vocabulary and domain knowledge to listening comprehension.

Present study

In the present study, we asked the following two research questions: (1) Are there distinct profiles of students who show similar patterns of performance on measures of vocabulary and domain knowledge in the fall of kindergarten? (2) Is profile membership associated with listening comprehension development during the kindergarten year? As this is the first study, to our knowledge, to investigate profiles of students' vocabulary and domain knowledge, the number of groups which could be expected from latent profile analysis is unknown. We hypothesised that there would be at least one group of students whose achievement in domain knowledge (science and/or social studies) is lower than their achievement in vocabulary (receptive and/or expressive) because instruction for science and social studies often increases after entering elementary school. In addition, it is

plausible that there are different groups of students who demonstrate distinctive performance on receptive and expressive vocabulary measures, as well as science and social studies domain knowledge measures, due to different instructional support received during pre-kindergarten. Membership in profiles of vocabulary and domain knowledge is likely to predict listening comprehension development, based on the theory of comprehension (Kintsch, 2013). In sum, this work provides a nuanced understanding of how vocabulary and domain knowledge manifest in the beginning of kindergarten and relate to subsequent listening comprehension development during the kindergarten year.

Method

Participants

Participants were 629 kindergarten students (306 boys and 323 girls) from two large urban districts (located in Mid-Atlantic and Southern states) who were part of a larger project which encompassed two randomised controlled trials (Cabell & Hwang, 2020). Of students for whom parental permission was provided, approximately 30 students per school were randomly selected, evenly distributed among 70 kindergarten classrooms in 23 schools. There were no exclusionary criteria for participation. All students were part of the control group in the larger project and consequently experienced business-as-usual literacy practices in their respective districts. Students had a mean age of 67.7 months ($SD = 3.84$) at the start of the study.

Percentage of students with minoritised status (non-White) was 95.7% and 60% in the Mid-Atlantic and Southern districts, respectively. Mid-Atlantic and Southern districts were 88.5% and 40% Black; 5.1% and 25% Latinx; 0.8% and 4% Asian; 1.3% and 0% other races, respectively. Approximately 83% of students were eligible for free or reduced lunch service, and 4% of students were English learners in the Mid-Atlantic district; 44% students were eligible for the lunch service and 20% students were English learners in the Southern district. Across the two districts, approximately 13% of students received special education services.

General procedures and measures

Data were collected during the fall and spring of kindergarten in the 2017–2018 and 2018–2019 school years in the Mid-Atlantic and Southern urban school districts in the United States, respectively. Identical procedures were employed in each district. During 7- to 8-week testing windows in the fall and spring of kindergarten, students were individually administered a broad battery of assessments of approximately 45 minutes in length. Assessments took place in a quiet place in the school building. Prior to collecting individual student data, assessors completed a standardised training that typically included: (1) viewing an online training, (2) passing a quiz, (3) attending an in-person training session, (3) practicing the assessments, (4) observing administration by an experienced assessor and (5) being observed in the field.

Measures used in the present analyses represent a subset of the full battery used in the larger project. Students were assessed on measures of vocabulary and domain knowledge in the fall of kindergarten and on measures of listening comprehension in both fall and spring of kindergarten.

Vocabulary. Students were assessed on their receptive vocabulary knowledge using the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-IV; Dunn & Dunn, 2007). For each item, the administrator presents four illustrations, and the student must identify the picture that best represents the word spoken by the assessor. Internal consistency reliability for the measure is reported between at about .96. Students' expressive vocabulary was measured using the Picture Vocabulary subtest of the *Woodcock-Johnson III Tests of Achievement* (WJ III ACH; Woodcock et al., 2001). The administrator asks students to name objects depicted in a series of pictures. This subtest has acceptable reliability and validity with a median split-half reliability of .81.

Domain knowledge. Domain knowledge was measured using two of the Academic Knowledge subtests of the WJ III ACH: Science and Social Studies. The first section of each subtest requires response by pointing and the remainder of the items requires oral response. The Science subtest assesses information related to biological and physical sciences; the Social Studies subtest assesses information related to history, geography, government and economics (internal consistency = .88).

Listening comprehension. Listening comprehension skills (i.e., sentence and story comprehension) were measured with the Sentence Structure subtest of the Clinical Evaluation of Language Fundamentals-Fourth Edition (CELF-SS; Semel et al., 2004) and the narrative comprehension portion of the Test of Narrative Language (TNL; Gillam & Pearson, 2004). The CELF-SS assessed students' ability to understand increasingly complex sentences by selecting the picture that most closely represents the sentence provided by the assessor. Internal consistency reliability estimates average .87 across ages. The TNL assesses students' ability to understand stories. Students were assessed individually on the ability to comprehend stories in three formats: with no picture cues, with five sequenced pictures and with a single picture. For each of the three stories, the assessor read the story and then asked the student literal and inferential comprehension questions ($\alpha = .87$). The student relied on their memory to answer questions about the story without picture cues. For the stories with the picture cue, the assessor showed the student the picture cue that illustrates the story, while reading the story. The student was allowed to see the picture cue to answer comprehension questions.

Analytic approach

We followed the manual maximum likelihood (ML) approach to run latent profile analysis (LPA) as suggested by Nylund-Gibson et al. (2019). This approach can produce unbiased results by accounting for profile classification errors in estimating the relations between latent profile variables and outcome variables. Also, it can flexibly estimate the associations of covariates with latent profile variables. We did not use the conventional classify-analyse two steps (see p. 971 in Nylund-Gibson et al., 2019, for the description) to avoid biased estimates due to classification errors (Vermunt, 2010).

In the first step, we explored multiple unconditional models that classify students into different numbers of groups, based on their scores across two vocabulary and two domain knowledge measures in the fall of kindergarten. As there has been no prior study examining the profiles of vocabulary and domain knowledge, we had to explore and compare multiple models to determine the best fitting model (i.e., the optimal number of

groups; Nylund et al., 2007). The vocabulary and comprehension measures were *z*-scored in order to place all variables on the same scale. Using *Mplus* 8 (Muthén & Muthén, 2015), we fitted and explored different models by increasing the number of groups from two through five and then selecting the best fitting unconditional model (i.e., the number of groups between two and five) based on multiple indices (see Logan & Pentimonti, 2016, for description of indices), following guidelines in the extant literature (e.g., Nylund et al., 2007). Specifically, we used Bayesian information criterion (BIC), Akaike information criterion (AIC), Lo–Mendell–Rubin likelihood ratio test bootstrap test (LLRT; Lo et al., 2001) and entropy value (Ramaswamy et al., 1993). Poster probabilities and group assignment for the selected model were saved into a separate data file for the following steps. Classification errors were automatically computed by the software and used in the next step to predict listening comprehension outcomes.

In the second step, we examined sentence and story comprehension development throughout kindergarten for each group identified in the first step. To estimate the means of the residualised change in listening comprehension from the fall to spring for each group, fall comprehension scores were included in the analysis as covariates (e.g., Dickinson et al., 2019) along with age and gender, while spring comprehension measures were included as outcomes. Differences of each group mean of residualised changes in listening comprehension measures were tested by using model constraint commands (see p. 982 in Nylund-Gibson et al., 2019, for an example of the *Mplus* syntax). Classification errors (obtained from the first step) were included in the model as fixed values between the latent group variable and the group membership variable. For both steps, a sandwich estimator was used to adjust the standard errors for the intraclass correlation coefficient, as student-level data were nested within school-level data.

Missing values

Outliers in each measure were recoded as missing (e.g., Hart et al., 2016) because extreme scores can distort means of latent groups and significance of BLRT (Nylund et al., 2007). Outliers were scores smaller than -3 standard deviations (SD) or larger than $+3$ SDs from the mean. A sensitivity analysis was conducted by removing scores smaller than -2 SDs or greater than $+2$ SDs from the mean. As the same results were observed, the analysis without three-standard-deviation outliers were reported.

A Missing Completely at Random (MCAR) test (Little & Rubin, 1987) indicated that the data were not missing completely at random ($\chi^2[318] = 381.30, p < .01$). Out of 629 students, 534 students did not have any missing values. Pretest measures had approximately 1% through 3% of data missing, and the listening comprehension measures in the spring of kindergarten had roughly 10% of data missing. Missing data were handled using the full information maximum likelihood (FIML) estimator (Muthén & Muthén, 2015).

Results

Descriptive statistics and correlations are summarised in Table 1. Strong correlations (approximately between .6 and .7) were observed between the vocabulary and listening comprehension measures (cf. $r = .31$ and $.41$ in Kim, 2020). Also, similar coefficients were observed in correlations of the domain knowledge measures (WJ III ACH science and social studies) with the listening comprehension (cf. $r = .3$ for science and reading

Table 1. Descriptive statistics and correlation matrix

Variable	<i>n</i>	M		1	2	3	4	5	6	7
		SD	Max							
Vocabulary (Fall)	1. Receptive	617	83.14	17						
			22.91	145						
	2. Expressive	614	15.30	3	.73					
			4.07	27						
Domain knowledge (Fall)	3. Science	613	10.70	4	.68	.65				
			2.10	17						
	4. Social Studies	617	9.89	2	.69	.68	.62			
			2.58	16						
Listening Comprehension (Fall)	5. Sentence	623	16.32	3	.65	.56	.55	.60		
			4.61	26						
	6. Narrative	621	15.02	0	.65	.58	.59	.67	.67	
			7.63	35						
Listening Comprehension (Spring)	7. Sentence	571	19.96	7	.64	.58	.53	.62	.66	.61
			4.20	26						
	8. Narrative	551	20.20	4	.64	.59	.58	.67	.59	.72
			6.55	36						.66

Note: All correlation coefficients were significant at the level of .001. Raw scores were used for each measure. Min = minimum, Max = maximum.

comprehension in Hwang & Duke, 2020) and vocabulary measures (cf. $r = .39$ between topic knowledge and vocabulary in Rydland et al., 2012). Standard scores were also calculated for the receptive (PPVT-IV) vocabulary measure in the fall of kindergarten and the story comprehension (TNL) measures in the fall and spring of kindergarten. On average, students scored as well as or better at the beginning of kindergarten than 40% of the normative sample for the receptive vocabulary measure. For the story comprehension measure, students scored as well as or better than 25% of the normative sample in the fall semester and 34% in the spring semester.

To answer the first research question, the profiles based on students' (receptive and expressive) vocabulary and (science and social studies) domain knowledge were explored. Four models, each with a different number of groups (from two through five), were developed, and their fit indices were compared with one another (see Table 2). Based on AIC and BIC indices, the five-group model was the best fitting model as it showed the lowest values across both indices. However, as illustrated in Figure 1, the slopes begin to flatten after the three-group model, meaning very little information is gained compared with the number of degrees of freedom used to estimate models with more groups (i.e., four-group and five-group models). Moreover, entropy and LLRT indicated that the three-group model was the best fitting model. Highest entropy value was observed with the three-group model, and a series of LLRT were statistically significant until the four-group model. Balancing these results, the three-group model was selected.

Figure 2 presents the results from the three-group model. The first group extracted from LPA had the smallest number of kindergarteners among the three groups ($n = 89$; 14%) and

Table 2. Model fit indices for each tested model

Model	AIC	BIC	Entropy	LLRT
2 groups	6156.996	6214.666	0.840	0.0001
3 groups	5741.291	5821.142	0.842	0.0051
4 groups	5640.085	5742.117	0.815	0.0982
5 groups	5582.500	5706.712	0.806	0.4839

Note: LLRT indicates *p*-values from Lo–Mendell–Rubin likelihood ratio test.
Bold indicates best fitting model selected.

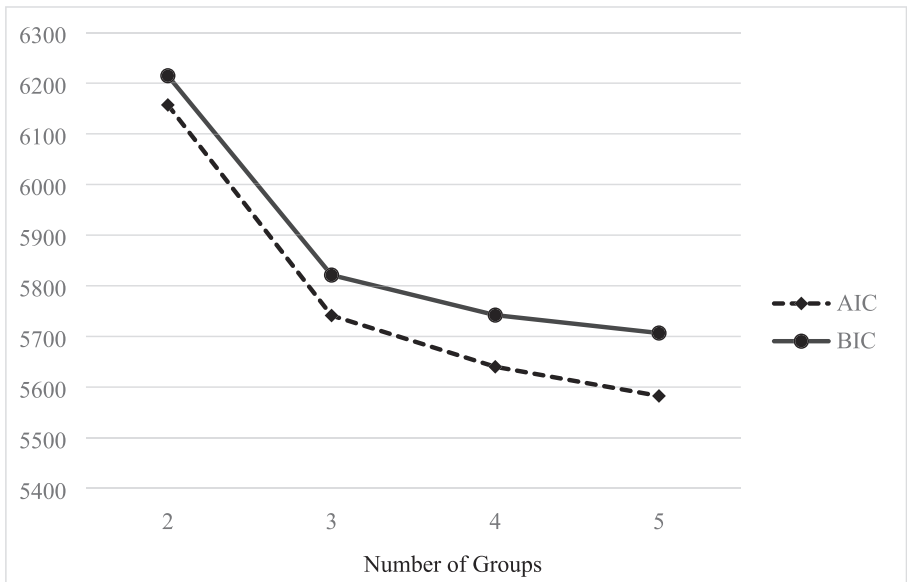
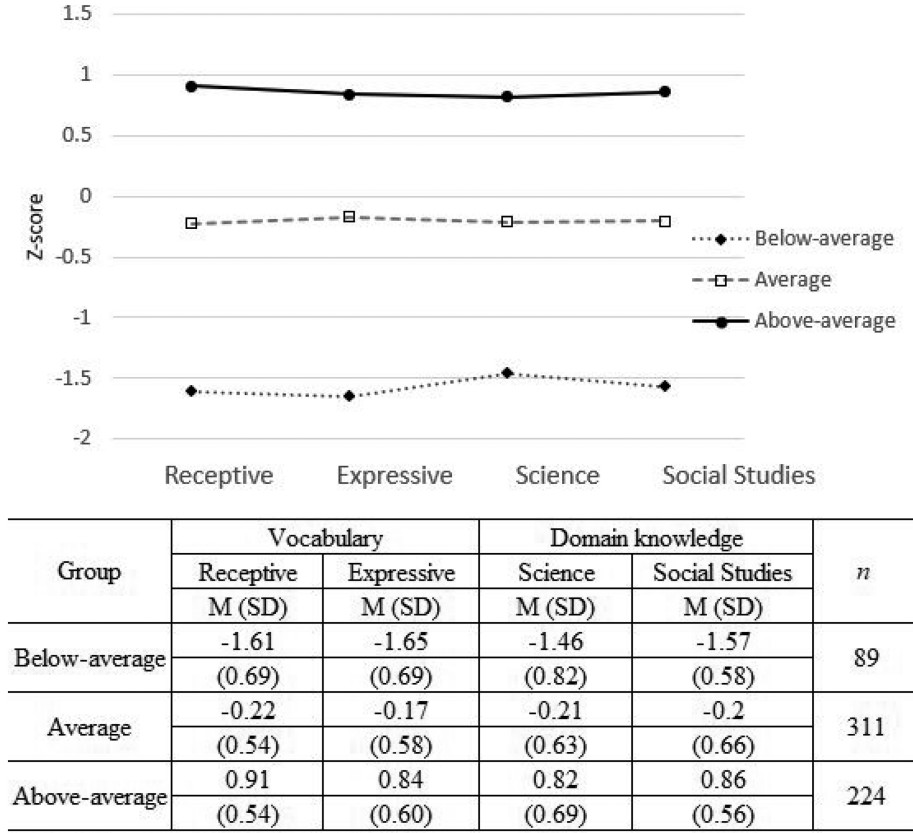


Figure 1. Trends in AIC and BIC across latent profiles.

was characterised by receptive and expressive vocabulary achievement approximately 1.6 SD below the mean obtained from the sample, with similarly low achievement in science and social studies domain knowledge (hence after below-average group). The correlations among the vocabulary and domain knowledge measures ranged between .18 and .39. All correlations were significant except the correlation between science domain knowledge and receptive vocabulary.

In addition, half of kindergartners ($n = 311$; 50%) were observed to belong to the second group. This group of students showed scores of the receptive and expressive vocabulary measures and science and social studies domain measures slightly below the sample mean (-0.2 SD; hence after average group). Within the average group, the correlations of receptive vocabulary with expressive vocabulary and science domain knowledge were statistically significant, of which coefficients were approximately .15. However, receptive vocabulary was not significantly correlated with social studies domain knowledge ($r = .11$). Expressive vocabulary was not significantly correlated with the two domain knowledge measures ($r \approx .09$). Also, the correlation between science and social studies domain knowledge measures was not significant ($r = -.003$).



Note . M = mean, SD = standard deviation

Figure 2. Results from the selected model, including the sample size of each group and descriptives of z-scores of the vocabulary and domain knowledge measures.

The last group of students ($n = 224$; 36%) displayed the highest scores of the vocabulary and domain knowledge measures (approximately 0.8 SD above the sample mean; hence after above-average group). All correlation coefficients among the vocabulary and domain knowledge measures were statistically significant ($.17 < r < .25$), except the correlation between receptive vocabulary and social studies domain knowledge ($r = .13$).

A series of ANOVAs were conducted to determine whether performance on the vocabulary and domain knowledge measures was statistically different across the three groups. The results of ANOVAs showed that group membership was significantly related to receptive vocabulary ($F[2, 614] = 664.2, p < .001$), expressive vocabulary ($F[2, 611] = 532.11, p < .001$), science domain knowledge ($F[2, 610] = 362.87, p < .001$) and social studies domain knowledge ($F[2, 614] = 513.71, p < .001$). Tukey's post hoc tests indicated that all pair-wise comparisons among the three groups were statistically significant ($p < .001$).

To address the second research question, a series of chi-square test were conducted to test whether changes in the listening comprehension outcomes between the fall and spring semesters were statistically different across the three groups. Table 3 summarises group-specific means of residualised changes and significance tests of pair-wise comparisons for residualised changes by group. As expected, the below-average group in terms

Table 3. Group-specific means of both distal outcomes of listening comprehension

Listening Comprehension ¹	Group ²	Distal mean	Pairwise comparison	
			Average	Above average
Sentence	Below average	14.89	−4.76***	−7.07***
	Average	19.65		−2.32***
	Above average	21.97		
Narrative	Below average	13.29	−5.55***	−9.70***
	Average	18.83		−4.16***
	Above average	22.99		

Note: Superscript (1) indicates in the winter of kindergarten, whereas superscript (2) indicates grouping based on vocabulary and domain knowledge measures in the fall of kindergarten.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

of vocabulary and domain knowledge in the fall semester displayed the lowest residualised gains of the listening comprehension measures in the spring semester, compared with the average and above-average groups (see third column in Table 3); the above-average group showed the highest residualised gains on the listening comprehension measures. Significance tests indicated that all pair-wise comparisons were statistically significant at .001 level (see fourth and fifth columns in Table 3). Effect sizes (Hedges' g) were calculated for residualised gains of both listening comprehension measures between each pair of groups. Overall, effect sizes were slightly larger on sentence comprehension than those on story comprehension residualised gains. The effect size between the below-average and average groups was 1.07 on sentence comprehension and 0.64 on story comprehension residualised gains; between the below-average and above-average groups, 1.86 on sentence comprehension and 1.12 on story comprehension residualised gains. Moderate effect sizes were observed between the average and above-average groups, 0.65 on sentence comprehension and 0.51 on story comprehension residualised gains.

Discussion

Given the importance of vocabulary and domain knowledge in listening comprehension development, the current study explored whether there are distinct profiles among kindergarteners who have different strengths and weaknesses in receptive and expressive vocabulary and science and social studies domain knowledge (the fall of kindergarten) and investigated the relations between the profiles in the fall of kindergarten and residualised gains in listening comprehension between the fall and spring of kindergarten. Our findings indicate that although there were distinct profiles, students started school with commensurate achievement in receptive and expressive vocabulary and science and social studies domain knowledge, and these developmental patterns of performance appeared to be meaningful to listening comprehension development throughout kindergarten year.

Commensurate achievement of vocabulary and domain knowledge

The results of latent profile analysis showed three groups of students who had commensurate achievement in vocabulary and domain knowledge, namely, students who performed below average, average, and above average on the vocabulary and domain knowledge measures relative to their peers. (Note that average here is the mean scores in the sample in this study, not the normative sample.) There was no group of students who demonstrated unbalanced achievement in vocabulary and domain knowledge (e.g., relatively high vocabulary with relatively low domain knowledge or vice versa).

The comparable achievement between vocabulary and domain knowledge might be partially attributable to mutually enhancing relations between acquisition of vocabulary and domain knowledge. The process of learning vocabulary is guided by conceptual information students have learned (Borovsky et al., 2016). Conceptual information classifies objects (e.g., alpacas) based on their features and elements (e.g., eating hay and grass) into superordinate categories (e.g., alpacas belong to herbivores; Fitzgerald et al., 2017). Students with strong domain knowledge are likely to have more conceptual information, by which acquisition of vocabulary is facilitated to the similar extent with domain knowledge. The process of learning new words in students with weak domain knowledge might be inefficient due to less conceptual information available to them, thus resulting in a weak vocabulary commensurate with their domain knowledge. On the other hand, using vocabulary to refer to or describe objects can support learning of conceptual information because it facilitates the categorisation process (Lupyan et al., 2007). The extent to which students know words are likely to influence how frequently they use words. The use of vocabulary, in turn, can support students to gain conceptual information and subsequently cultivate domain knowledge to the similar extent with vocabulary.

In addition, the similar performance on domain knowledge and vocabulary might be partially explained by how domain knowledge was measured. The standardised measures for science and social studies domain knowledge were administered via oral language. All of test items were presented orally, and students were requested to provide verbal answers for most of the items. Thus, students might have encountered difficulties in understanding and responding to the test items even though they might have had nonlinguistic mental representations related to the test items (e.g., visual images). With measures that can assess domain knowledge with alternative forms (e.g., drawing), the performance on domain knowledge might not be commensurate with that of vocabulary.

Looking closely, similar scores were also observed between receptive and expressive vocabulary. That is, students at the beginning of schooling recognised words to a similar extent that they verbally produced words. These findings are in line with Cabell et al. (2011) and Cabell et al. (2013) who found commensurate achievement in receptive and expressive vocabulary in language and literacy profiles of children at both the beginning and the end of pre-kindergarten. Support for vocabulary at home and preschool might have led to commensurable achievement between receptive and expressive vocabulary. For example, read-alouds might enhance receptive and expressive vocabulary to the extent that they are commensurate with each other, even though characteristics of read-alouds can be differentially related to receptive and expressive vocabulary (Gonzalez et al., 2014).

In addition, comparable achievement was observed between science and social studies domain knowledge. Even though the two domains are considered distinct from one another, students arrived at school with similar amounts of domain knowledge in science and social studies. The reason for this finding is not clear. It may be that parents (along with

television programming and other learning activities) tend to expose their children to general knowledge that cuts across domains of both science and social studies, rather than reinforcing one over the other. Alternatively, this finding might be related to fact that the measures did not assess discipline-specific ways of knowing (e.g., drawing timelines for history and using a microscope for science), and the items asked were more related to content students might have encountered daily interaction with adults, books and media.

Relation of the profiles with listening comprehension development

Chi-square tests revealed that three groups identified based on vocabulary and domain knowledge (fall of kindergarten) displayed statistically different residualised gains in listening comprehension between the fall and spring of kindergarten. The differences among the groups were observed in both sentence and story comprehension measures, which indicates that the differences in vocabulary and domain knowledge are related to sentence-level and discourse-level listening comprehension development. The findings of the study are in line with previous research that has demonstrated the positive role of vocabulary (Wright & Cervetti, 2016) and domain knowledge (Hwang, 2019) in reading comprehension, while extending the literature by focusing on listening comprehension and identifying profiles of students regarding vocabulary and domain knowledge.

Limitations of the study

The current study has several limitations. First, as discussed earlier, the standardised measures for domain knowledge were administered via oral language, and accordingly, knowledge stored as nonlinguistic forms (e.g., images, sounds) in a long-term memory was not considered. A further study is recommended to develop and use alternative measures to assess domain knowledge via nonlinguistic formats (e.g., drawing, demonstrating how to use experiment devices) in combination with traditional measures such as the ones in the current study. Second, the current study was not able to consider different genres of listening comprehension as the listening comprehension measures contain story passages or single sentences. Further investigation is recommended to examine the possibility of whether the relations between the profiles and listening comprehension might be different by text genres. Third, the findings of the current study were based on the data obtained from urban districts, which might limit the external validity of the findings. Similar studies can be conducted with different populations from various regions to fully understand profiles of vocabulary and domain knowledge and their role in listening comprehension. Fourth, the results might have been influenced to some extent by presence of a third variable, general intelligence, that was not included in the study. Yet general intelligence would not account for the entire results of the study, based on literature that indicates the importance of the variables in the study even when general intelligence is considered (National Early Literacy Panel, 2008; Schneider et al., 1989).

Conclusion and implications for practice

Acknowledging the limitations, the current study is one of first that has revealed how individual differences across receptive and expressive vocabulary and science and social studies domain knowledge manifest in students at the beginning of schooling.

Commensurate achievement in vocabulary and domain knowledge was observed within each of the three groups (low-average, average and above-average). Considerable achievement gaps across the groups were found in the residualised gains of listening comprehension between fall and spring of kindergarten, indicating the importance of vocabulary and domain knowledge in supporting listening comprehension development.

The commensurate achievement in vocabulary and domain knowledge appears to indicate that students have acquired vocabulary knowledge to the extent that they have developed domain knowledge and vice versa. Therefore, literacy instruction to foster vocabulary and domain knowledge simultaneously might be more supportive of listening comprehension than instruction focusing on either vocabulary or domain knowledge alone. Moreover, literacy instruction that supports domain knowledge can serve as a meaningful context in which to teach vocabulary because word meanings are taught in connection with concepts (Cabell & Hwang, 2020). For example, Neuman and Kaefer (2018) demonstrated that literacy instruction can support both science domain knowledge and vocabulary in kindergarten by reading aloud conceptually connected texts and explicitly teaching word meanings and categorical relations among words during the content-focused read-alouds.

Acknowledgements

We thank Dr. James S. Kim and Dr. Thomas G. White for their conceptual contributions to the larger project from which these data were drawn. We also thank Marcy Wyatt, Natalie Maus, Ansley Kramer and Christine Raynor for their contributions to management and execution of this project. Finally, we are grateful for the many administrators, teachers, students and parents who made this research possible. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A170635 awarded to Florida State University (PI: Cabell). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Data availability statement

Research data are not shared.

References

- Anderson, R.C. & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. International Reading Association.
- Borovsky, A., Ellis, E.M., Evans, J.L. & Elman, J.L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in two-year-olds. *Developmental Science*, 19(6), 918–932. <https://doi.org/10.1111/desc.12343>. Lexical
- Brandão, A.C.P. & Oakhill, J. (2005). “How do you know this answer?” – Children’s Children’s use of text data and general knowledge in story comprehension. *Reading and Writing*, 18(7–9), 687–713. <https://doi.org/10.1007/s11145-005-5600-x>
- Cabell, S.Q. & Hwang, H. (2020). Building content knowledge to boost comprehension in the primary grades. *Reading Research Quarterly*, 55(S1), 1–9. <https://doi.org/10.1002/trq.338>

- Cabell, S.Q., Justice, L.M., Konold, T.R. & McGinty, A.S. (2011). Profiles of emergent literacy skills among preschool children who are at risk for academic difficulties. *Early Childhood Research Quarterly*, 26(1), 1–14. <https://doi.org/10.1016/j.ecresq.2010.05.003>
- Cabell, S.Q., Justice, L.M., Logan, J.A.R. & Konold, T.R. (2013). Emergent literacy profiles among prekindergarten children from low-SES backgrounds: Longitudinal considerations. *Early Childhood Research Quarterly*, 28(3), 608–620. <https://doi.org/10.1016/j.ecresq.2013.03.007>
- Cain, K., Oakhill, J.V., Barnes, M.A. & Bryant, P.E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, 29(6), 850–859.
- Cervetti G. N., Wright T. S. (2020). The role of knowledge in understanding and learning from text. In EB Moje, P Afflerbach, P Enciso, NK Lesaux (Eds.), *Handbook of reading research: Volume V* (pp. 237–260). New York, NY: Routledge.
- Chien, N.C., Howes, C., Burchinal, M., Pianta, R.C., Ritchie, S., Bryant, D.M. et al. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development*, 81(5), 1534–1549. <https://www.jstor.org/stable/40800690>
- Cook, A. & Gueraud, S. (2005). What have we been missing? The role of general world knowledge in discourse processing. *Discourse Processes*, 39(2), 265–278. https://doi.org/10.1207/s15326950dp3902%263_9
- Dickinson, D.K., Nesbitt, K.T., Collins, M.F., Hadley, E.B., Newman, K., Rivera, B.L. et al. (2019). Teaching for breadth and depth of vocabulary knowledge: Learning from explicit and implicit instruction and the storybook texts. *Early Childhood Research Quarterly*, 47, 341–356. <https://doi.org/10.1016/j.ecresq.2018.07.012>
- Duke, N.K. & Carlisle, J. (2011). The development of comprehension. In M.L. Kamil, P.D. Pearson, E. Moje & P. Afflerbach (Eds.), *Handbook of reading research*. Routledge.
- Dunn, L. & Dunn, L. (2007). *Peabody picture vocabulary test fourth edition (PPVT-IV)*. NCS Pearson.
- Elbro, C. & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading*, 17(6), 435–452. <https://doi.org/10.1080/10888438.2013.774005>
- Fitzgerald, W.J., Elmore, J., Kung, M. & Stenner, A.J. (2017). The conceptual complexity of vocabulary in elementary-grades core science program textbooks. *Reading Research Quarterly*, 52(4), 417–442. <https://doi.org/10.1002/rq.184>
- Gillam, R.B. & Pearson, N.A. (2004). *TNL: Test of narrative language*. Pro-ed.
- Gonzalez, J.E., Pollard-Durodola, S., Simmons, D.C., Taylor, A.B., Davis, M.J., Fogarty, M. et al. (2014). Enhancing preschool children's vocabulary: Effects of teacher talk before, during and after shared reading. *Early Childhood Research Quarterly*, 29(2), 214–226. <https://doi.org/10.1016/j.ecresq.2013.11.001>
- Gonzalez, J.E., Pollard-Durodola, S., Simmons, D.C., Taylor, A.B., Davis, M.J., Kim, M. et al. (2011). Developing low-income preschoolers' social studies and science vocabulary knowledge through content-focused shared book reading. *Journal of Research on Educational Effectiveness*, 4(1), 25–52. <https://doi.org/10.1080/19345747.2010.487927>
- Hagen, Å.M., Melby-Lervåg, M. & Lervåg, A. (2017). Improving language comprehension in preschool children with language difficulties: A cluster randomized trial. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 58(10), 1132–1140. <https://doi.org/10.1111/jcpp.12762>
- Hargrave, A.C. & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, 15(1), 75–90. [https://doi.org/10.1016/S0885-2006\(99\)00038-1](https://doi.org/10.1016/S0885-2006(99)00038-1)
- Hart, S.A., Logan, J.A.R., Thompson, L., Kovas, Y., McLoughlin, G. & Petrill, S. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology*, 108(2), 181–193. <https://doi.org/10.1037/edu0000045.A>
- Hogan, T.P., Adlof, S.M. & Alonzo, C.N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology*, 16(3), 199–207. <https://doi.org/10.3109/17549507.2014.904441>
- Hwang, H. (2019). The role of science domain knowledge and reading motivation in predicting informational and narrative reading comprehension in L1 and L2: An international study. *Learning and Individual Differences*, 76, 1–11. <https://doi.org/10.1016/j.lindif.2019.101782>
- Hwang, H. (2020). Early general knowledge predicts English reading growth in bilingual and monolingual students throughout the elementary years. *Elementary School Journal*, 121(1), 154–178. <https://doi.org/10.1086/709857>
- Hwang, H. & Duke, N.K. (2020). Content counts and motivation matters: Reading comprehension in third-grade students who are English learners. *AERA Open*, 6(1), 1–17. <https://doi.org/10.1177/2332858419899075>
- Kim, Y.G. (2020). Toward integrative reading science: The direct and indirect effects model of reading. *Journal of Learning Disabilities*, 53(6), 469–491. <https://doi.org/10.1177/0022219420908239>

- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kintsch, W. (2013). Revisiting the Construction-Integration model of text comprehension and its implications for instruction. In D.E. Alvermann, N.J. Unrau & R.B. Ruddell (Eds.), *Theoretical models and processes of reading*. International Reading Association.
- Little, R.J.A. & Rubin, D.B. (1987). *Statistical analysis with missing data*. John Wiley & Sons.
- Lo, Y., Mendell, N.R. & Rubin, D.B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778. <https://doi.org/10.1093/biomet/90.4.991>
- Logan, J. A. R., & Pentimonti, J. M. (2016). Introduction to latent class analysis for reading fluency research. In K.D. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications*. Springer: New York. https://doi.org/10.1007/978-1-4939-2803-3_11
- Lonigan, C.J. & Burgess, S.R. (2017). Dimensionality of reading skills with elementary-school-age children. *Scientific Studies of Reading*, 21(3), 239–253. <https://doi.org/10.1080/10888438.2017.1285918>
- Lonigan, C.J. & Whitehurst, G.J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, 13(2), 263–290. [https://doi.org/10.1016/S0885-2006\(99\)80038-6](https://doi.org/10.1016/S0885-2006(99)80038-6)
- Lupyan, G., Rakison, D.H. & McClelland, J.L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1083. <https://doi.org/10.1111/j.1467-9280.2007.02028.x>
- Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, 53(2), 253–279.
- Muthén, L.K. & Muthén, B.O. (2015). *Mplus User's Guide (No. 6; 7th ed.)*. Muthén & Muthén.
- Nation, K., Cocksey, J., Taylor, J.S.H. & Bishop, D.V.M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 51(9), 1031–1039. <https://doi.org/10.1111/j.1469-7610.2010.02254.x>
- National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. National Institute for Literacy. <https://lincs.ed.gov/publications/pdf/NELPSummary.pdf>
- Neuman, S.B. & Kaefer, T. (2018). Developing low-income children's vocabulary and content knowledge through a shared book reading program. *Contemporary Educational Psychology*, 52, 15–24. <https://doi.org/10.1016/j.cedpsych.2017.12.001>
- Neuman, S.B. & Roskos, K. (2012). Helping children become more knowledgeable through text. *Reading Teacher*, 66(3), 207–210. <https://doi.org/10.1002/TRTR.01118>
- Nylund, K.L., Asparouhov, T. & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>
- Nylund-Gibson, K., Grimm, R.P. & Masyn, K.E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling*, 26(6), 967–985. <https://doi.org/10.1080/10705511.2019.1590146>
- Ramaswamy, V., Desarbo, W.S., Reibstein, D.J. & Robinson, W.T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12(1), 103–124. <https://doi.org/10.1287/mksc.12.1.103>
- Richter, T., Isberner, M.-B., Naumann, J. & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, 17, 1–20. <https://doi.org/10.1080/10888438.2013.764879>
- Ricketts, J., Nation, K. & Bishop, D.V.M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, 11(3), 235–257. <https://doi.org/10.1080/10888430701344306>
- Rydland, V., Aukrust, V.G. & Fulland, H. (2012). How word decoding, vocabulary and prior topic knowledge predict reading comprehension. A study of language-minority students in Norwegian fifth grade classrooms. *Reading and Writing*, 25(2), 465–482. <https://doi.org/10.1007/s11455-010-9279-2>
- Schneider, W., Körkel, J. & Weinert, F.E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, 81(3), 306–312. <https://doi.org/10.1037/0022-0663.81.3.306>
- Semel, E., Wiig, E.H. & Secord, W.A. (2004). *Clinical evaluation of language fundamentals, fourth edition—Screening test (CELF-4 screening test)*. (4th edn). The Psychological Corporation/A Harcourt Assessment Company.
- Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, 24(1), 123–138.

- Sénéchal, M., Ouellette, G. & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D.K. Dickinson & S.B. Neuman (Eds.), *Handbook of early literacy research*. Guilford.
- Stahl, S.A., Hare, V.C., Sinatra, R. & Gregory, J.F. (1991). Defining the role of prior knowledge and vocabulary in reading comprehension: The retiring of number 41. *Journal of Reading Behavior*, 23(4), 487–508.
- Storch, S.A. & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38(6), 934–947. <https://doi.org/10.1037/0012-1649.38.6.934>
- Tarchi, C. (2010). Reading comprehension of informative texts in secondary school: A focus on direct and indirect effects of reader's prior knowledge. *Learning and Individual Differences*, 20(5), 415–420. <https://doi.org/10.1016/j.lindif.2010.04.002>
- Vermunt, J.K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4), 450–469. <https://doi.org/10.1093/pan/mpq025>
- Woodcock, R.W., McGrew, K.S. & Mather, N. (2001). *Woodcock–Johnson III*. Riverside.
- Wright, T.S. & Cervetti, G.N. (2016). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, 52(2), 1–24. <https://doi.org/10.1002/rtrq.163>

HyeJin Hwang is a Postdoctoral Scholar at the Florida Center for Reading Research in the Florida State University. Her research interests include content-area literacy in the elementary years, particularly for bilingual students.

Sonia Q. Cabell is an Assistant Professor at the College of Education and Florida Center for Reading Research in the Florida State University. Her research interests focus on the prevention of reading difficulties among young children who are at-risk, particularly those who are living in poverty.

Received 4 April 2020; revised version received 8 March 2021.

Address for correspondence: HyeJin Hwang, Florida Center for Reading Research, Florida State University, 2010 Levy Avenue Suite 100, Tallahassee, FL 32310, USA.
E-mail: hhwang4@fsu.edu